NTP
National Toxicology Program

Background on Draft NTP Approach for Systematic Review and Evidence Integration for Literature-Based Health Assessments

Andrew Rooney, Ph.D.
National Institute of Environmental Health Sciences

NTP Board of Scientific Counselors Meeting
December 11, 2012

Presentation Outline

- Background on systematic review
- Development of the draft NTP Approach
- The draft NTP Approach and evidence integration
- Specific aspects brought to the working group for comment
  - Step 4: Assessing the quality or risk of bias of individual studies
  - Step 5: Rating the confidence in the body of evidence
  - Step 6: Translating confidence ratings into evidence of health effects
  - Step 7: Integrating evidence to develop hazard identification conclusions
- Questions

Systematic Review

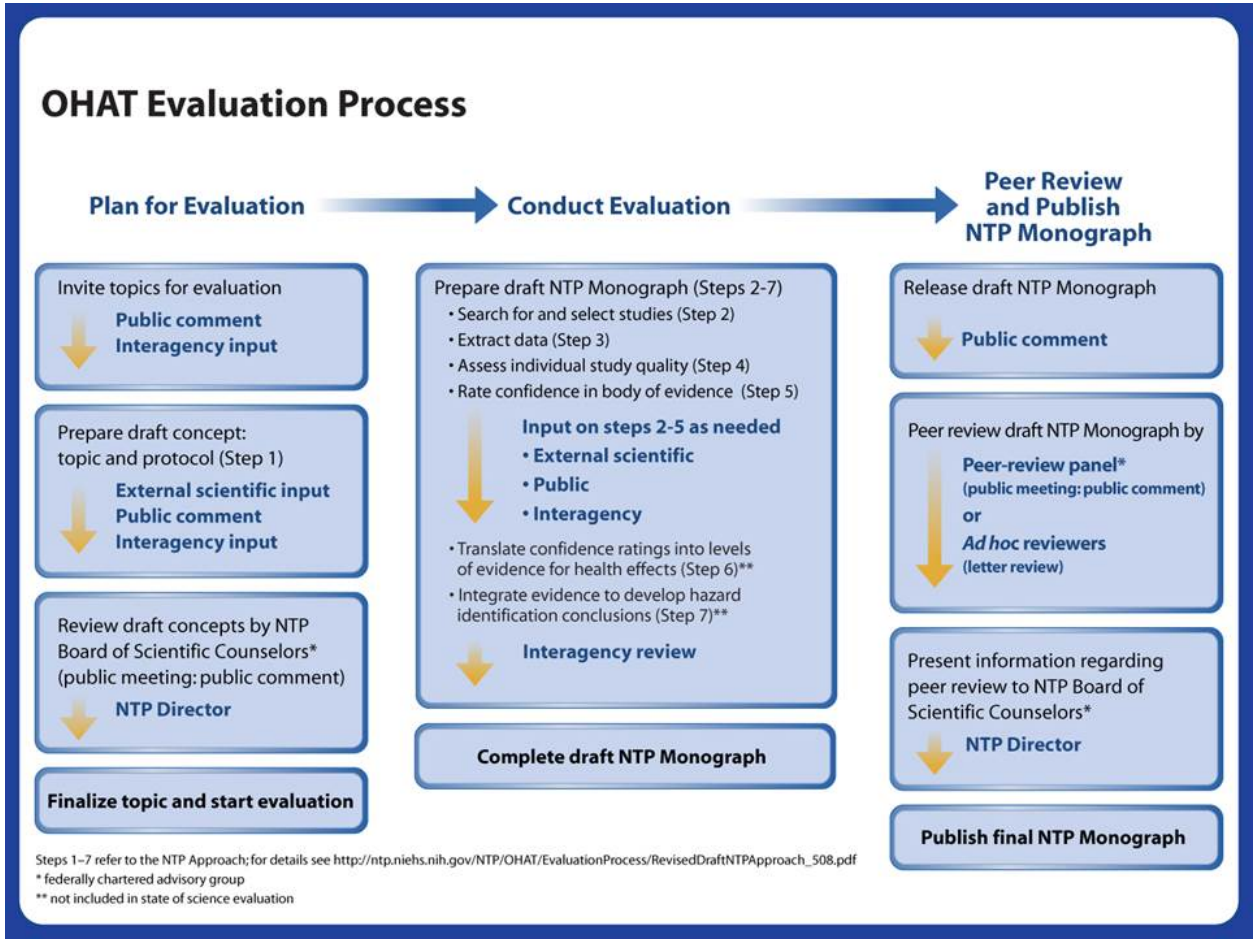
- A scientific investigation that focuses on a specific question, and uses explicit, pre-specified methods to identify, select, summarize, and assess the findings of similar studies
- Provides greater transparency
- Existing methods:
  - reach evidence-based conclusions
  - develop clinical or public health recommendations
  - clarify need for additional research
  - may or may not result in quantitative meta-analysis
- Existing methodologies are generally used for assessment of healthcare interventions

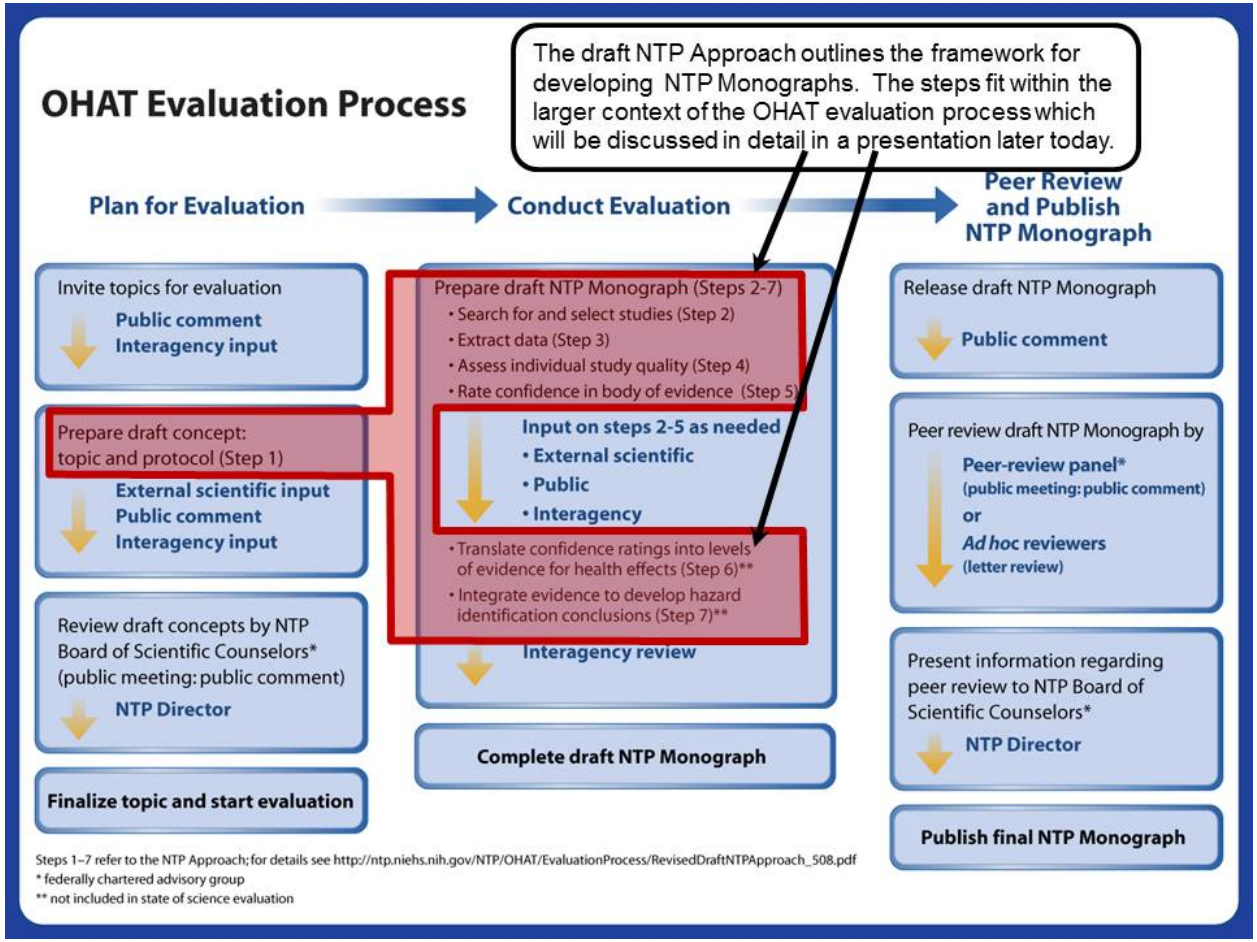
What Does A Systematic Review Not Do?

- Does not operate like an algorithm or computer program
- Does not eliminate the need for expert judgment
  - Systematic review provides a structure to document the basis of decisions
- Does not guarantee reproducibility of conclusions
  - Increased transparency does not necessarily eliminate differences in scientific judgment

Why Develop the NTP Approach?

- The NTP is adopting systematic review procedures for literature-based evaluations to enhance transparency for reaching and communicating health assessment conclusions
- Existing methods do not provide guidance on how to
  - Integrate evidence across human, animal, and mechanistic studies
  - Reach hazard identification conclusions


OHAT Evaluation Process
Plan for Evaluation
Conduct Evaluation
Peer Review and Publish NTP Monograph
Invite topics for evaluation
Public comment
Interagency input
Prepare draft concept: topic and protocol (Step 1)
External scientific input
Public comment
Interagency input
Review draft concepts by NTP Board of Scientific Counselors* (public meeting: public comment)
NTP Director
Finalize topic and start evaluation
Prepare draft NTP Monograph (Steps 2-7)
• Search for and select studies (Step 2)
• Extract data (Step 3)
• Assess individual study quality (Step 4)
• Rate confidence in body of evidence (Step 5)
Input on steps 2-5 as needed
• External scientific
• Public
• Interagency
• Translate confidence ratings into levels of evidence for health effects (Step 6)**
• Integrate evidence to develop hazard identification conclusions (Step 7)**
Interagency review
Complete draft NTP Monograph
Release draft NTP Monograph
Public comment
Peer review draft NTP Monograph by
Peer-review panel*
(public meeting: public comment)
or
Ad hoc reviewers
(letter review)
Present information regarding peer review to NTP Board of Scientific Counselors*
NTP Director
Publish final NTP Monograph
Steps 1–7 refer to the NTP Approach; for details see http://ntp.niehs.nih.gov/NTP/OHAT/EvaluationProcess/RevisedDraftNTPApproach_508.pdf
* federally chartered advisory group
** not included in state of science evaluation


OHAT Evaluation Process
The draft NTP Approach outlines the framework for developing NTP Monographs. The steps fit within the larger context of the OHAT evaluation process which will be discussed in detail in a presentation later today.
Plan for Evaluation
Conduct Evaluation
Peer Review and Publish NTP Monograph
Invite topics for evaluation
Public comment
Interagency input
Prepare draft concept: topic and protocol (Step 1)
External scientific input
Public comment
Interagency input
Review draft concepts by NTP Board of Scientific Counselors* (public meeting: public comment)
NTP Director
Finalize topic and start evaluation
Prepare draft NTP Monograph (Steps 2-7)
• Search for and select studies (Step 2)
• Extract data (Step 3)
• Assess individual study quality (Step 4)
• Rate confidence in body of evidence (Step 5)
Input on steps 2-5 as needed
• External scientific
• Public
• Interagency
• Translate confidence ratings into levels of evidence for health effects (Step 6)**
• Integrate evidence to develop hazard identification conclusions (Step 7)**
Interagency review
Complete draft NTP Monograph
Release draft NTP Monograph
Public comment
Peer review draft NTP Monograph by
Peer-review panel*
(public meeting: public comment)
or
Ad hoc reviewers
(letter review)
Present information regarding peer review to NTP Board of Scientific Counselors*
NTP Director
Publish final NTP Monograph
Steps 1–7 refer to the NTP Approach; for details see http://ntp.niehs.nih.gov/NTP/OHAT/EvaluationProcess/RevisedDraftNTPApproach_508.pdf
* federally chartered advisory group
** not included in state of science evaluation

Development of the Draft NTP Approach

- NTP systematic review webinars (Jan – May, 2012)
  - **Goal**: Increase understanding of issues relating to systematic review
  - **Format**: Expert and cross-agency discussions on concepts and existing methods

- Interagency communication
  - **Webinars**
    - **June 5**: "New Tools of Systematic Review, Information Management and Data Display"
    - **September 25**: "Systematic Review and New Tools of Information Management"
  - **NTP Executive Committee briefings**

- NTP Board of Scientific Counselors
  - **At the June 22 public meeting NTP staff outlined**
    - Background and advantages of systematic review to enhance transparency
    - OHAT development of tools for information management and data display
    - Plans to incorporate systematic review into NTP literature-based assessments. Plans included
      1) *Review of the NTP's Draft Approach by a NTP BSC Working Group in late summer of 2012*
      2) *Presentation of the Draft NTP Approach to the NTP BSC in December 2012 or Spring 2013*

Sources Considered

- Published systematic review methods and resources
  - **AHRQ** - Agency for Healthcare Research and Quality
  - **CAMARADES** - Collaborative Approach to Meta Analysis and Review of Animal Data from Experimental Studies
  - **Cochrane Collaboration**
  - **GRADE Working Group** - Grading of Recommendations, Assessment, Development, and Evaluation
  - **Navigation Guide Work Group**

- Technical expert consultation on concepts and existing methods
  - **Lisa Bero** - Director, Cochrane Center at UCSF
  - **Gordon Guyatt** - Co-chair, GRADE Working Group, McMaster University
  - **Malcolm Macleod** - CAMARADES Centre, University of Edinburgh
  - **Karen Robinson** - Co-Director, AHRQ Evidence-Based Practice Center, Johns Hopkins
  - **Holger Schünenmann** - Co-chair, GRADE Working Group, McMaster University
  - **Tracey Woodruff** - Director, Program on Reproductive Health and the Environment, UCSF

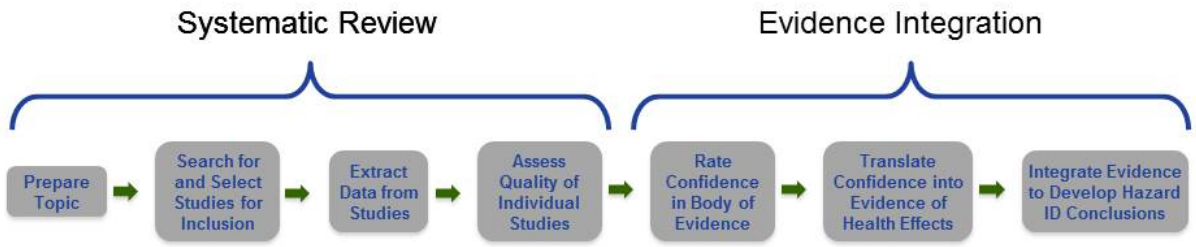
- NTP BSC Working Group to comment on draft NTP Approach

Presentation Outline

- Background on systematic review
- Development of the draft NTP Approach
- **The draft NTP Approach and evidence integration**
- Specific aspects brought to the working group for comment
  - Step 4: Assessing the quality or risk of bias of individual studies
  - Step 5: Rating the confidence in the body of evidence
  - Step 6: Translating confidence ratings into evidence of health effects
  - Step 7: Integrating evidence to develop hazard identification conclusions
- Questions

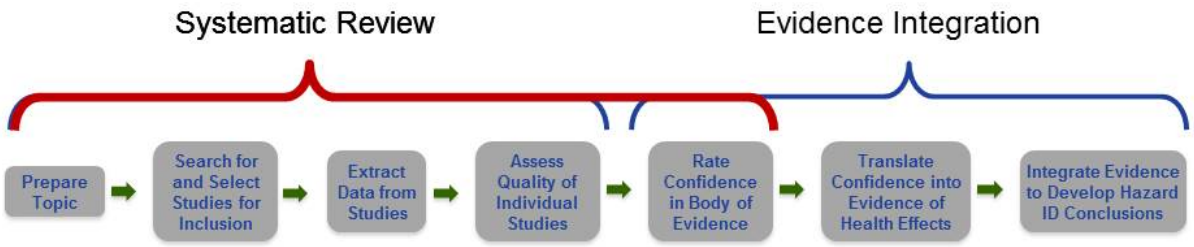
The Draft NTP Approach

- The NTP Approach builds on and extends existing methods for systematic review
- **Systematic review** is the basis for a transparent evaluation
- **Evidence integration** is the process of assessing and integrating the body of evidence to develop hazard ID conclusions



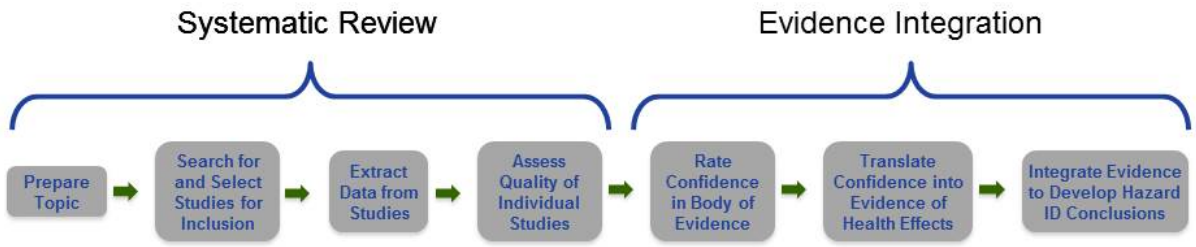
The Draft NTP Approach

- The NTP Approach builds on and extends existing methods for systematic review
- **Systematic review** is the basis for a transparent evaluation
- **Evidence integration** is the process of assessing and integrating the body of evidence to develop hazard ID conclusions



The Draft NTP Approach

- The NTP Approach builds on and extends existing methods for systematic review
- **Systematic review** is the basis for a transparent evaluation
- **Evidence integration** is the process of assessing and integrating the body of evidence to develop hazard ID conclusions



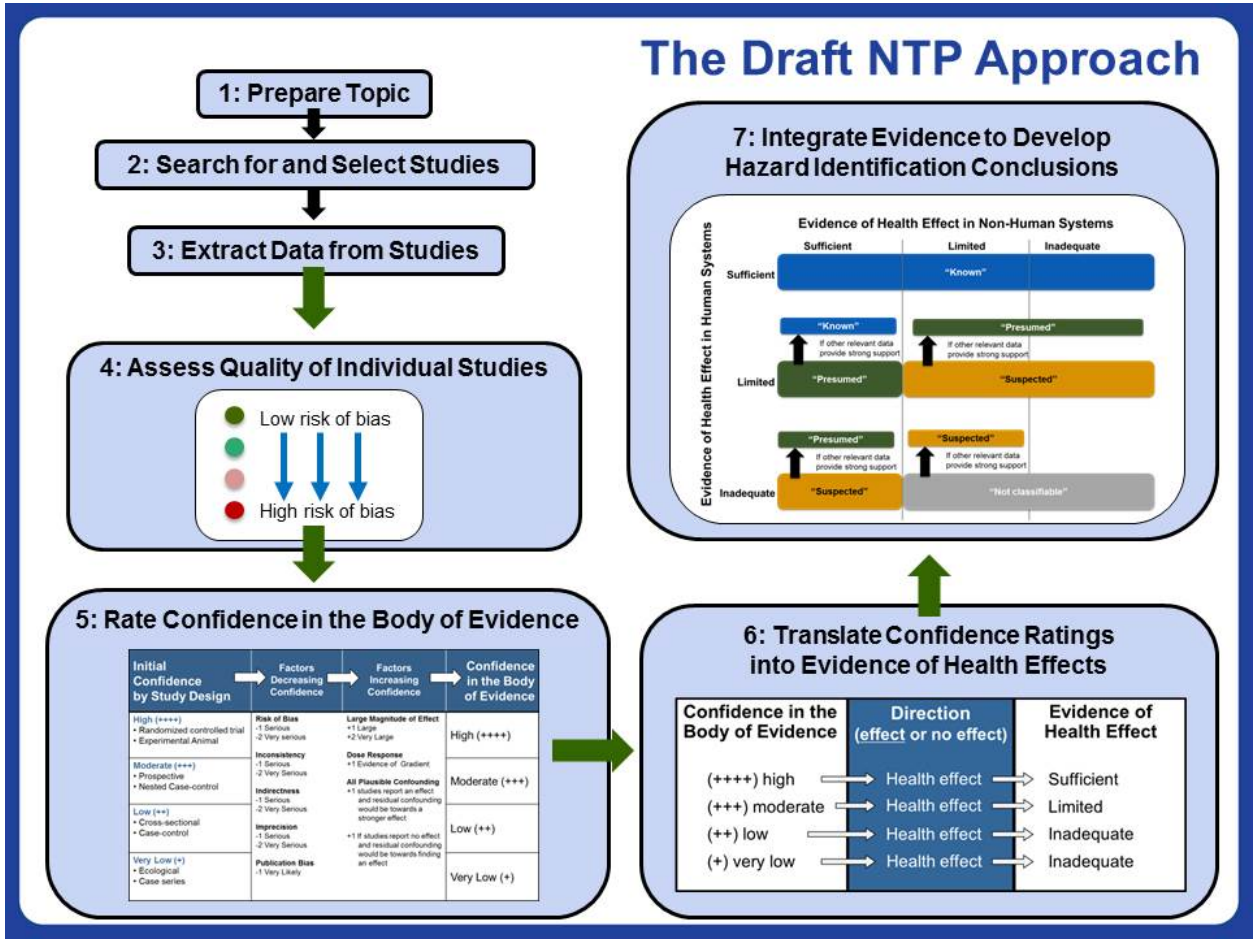
NTP
National Toxicology Program

What is Evidence Integration to the NTP?

- Evidence integration

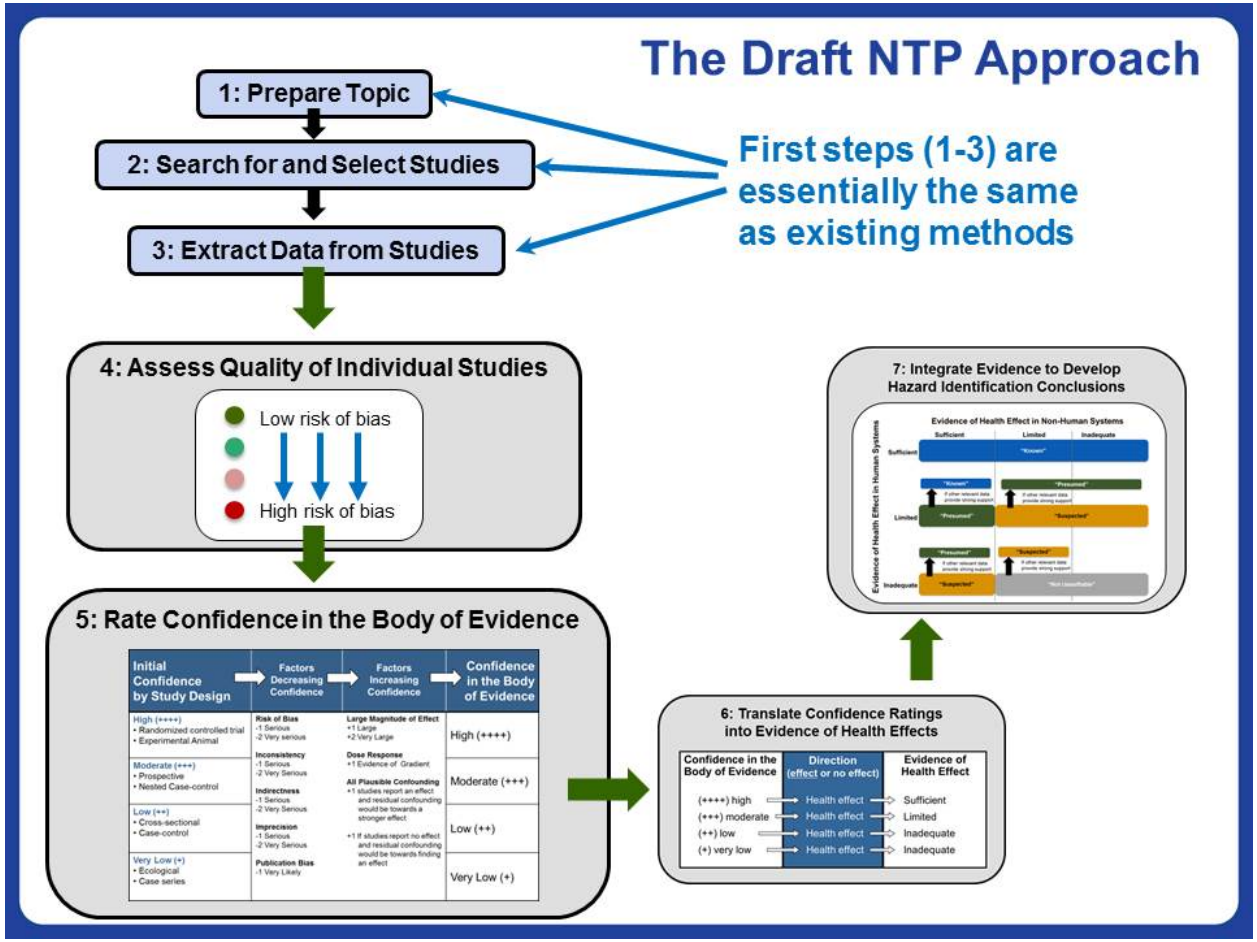
  process for reaching conclusions on the NTP's confidence across a body of studies within an evidence stream (i.e., human and animal data separately) and then integrating those conclusions across the evidence streams with consideration of other relevant data such as supporting evidence from mechanistic studies

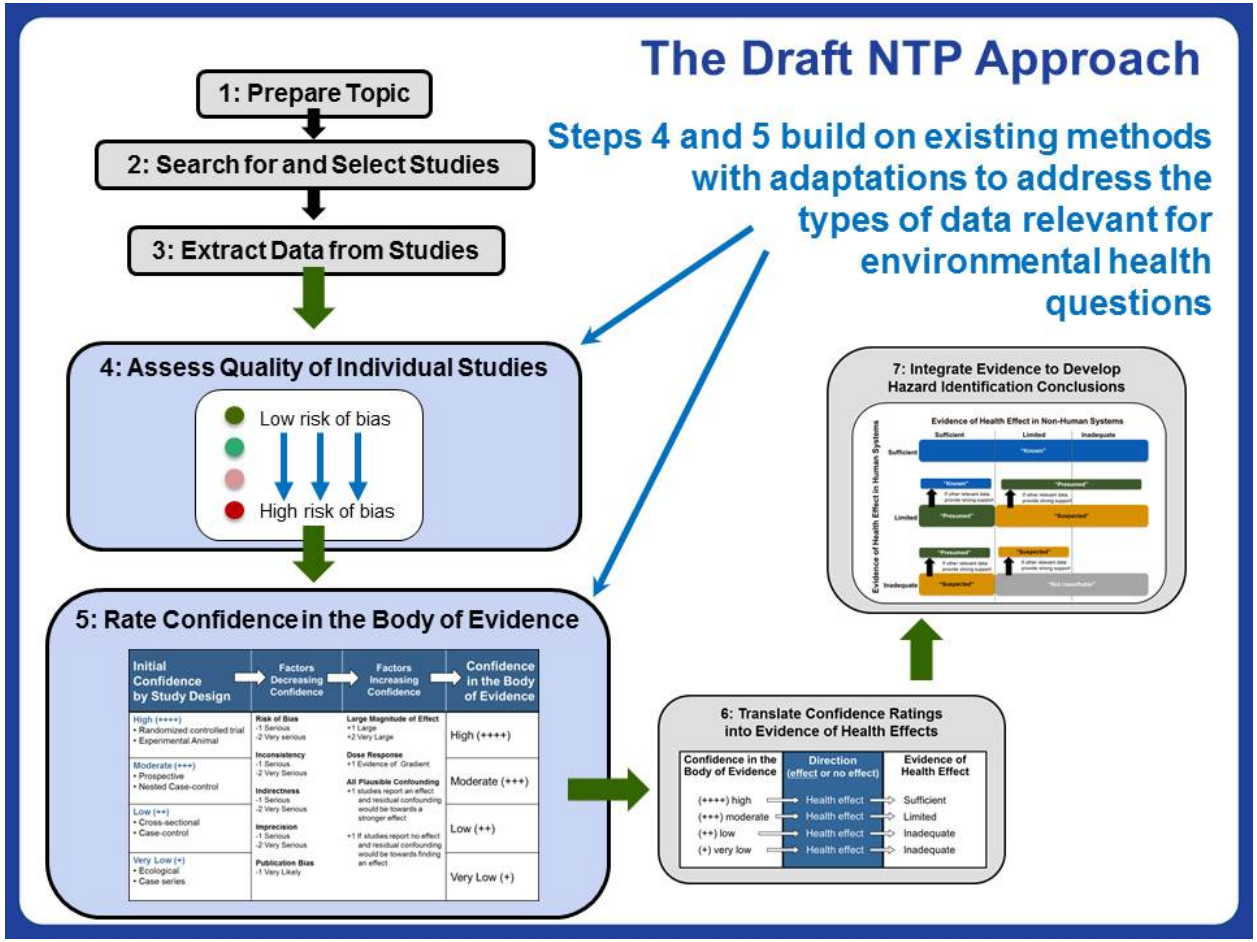
- Why not "Weight of Evidence"?
  - Lack of consensus on meaning (Weed et al., 2005)

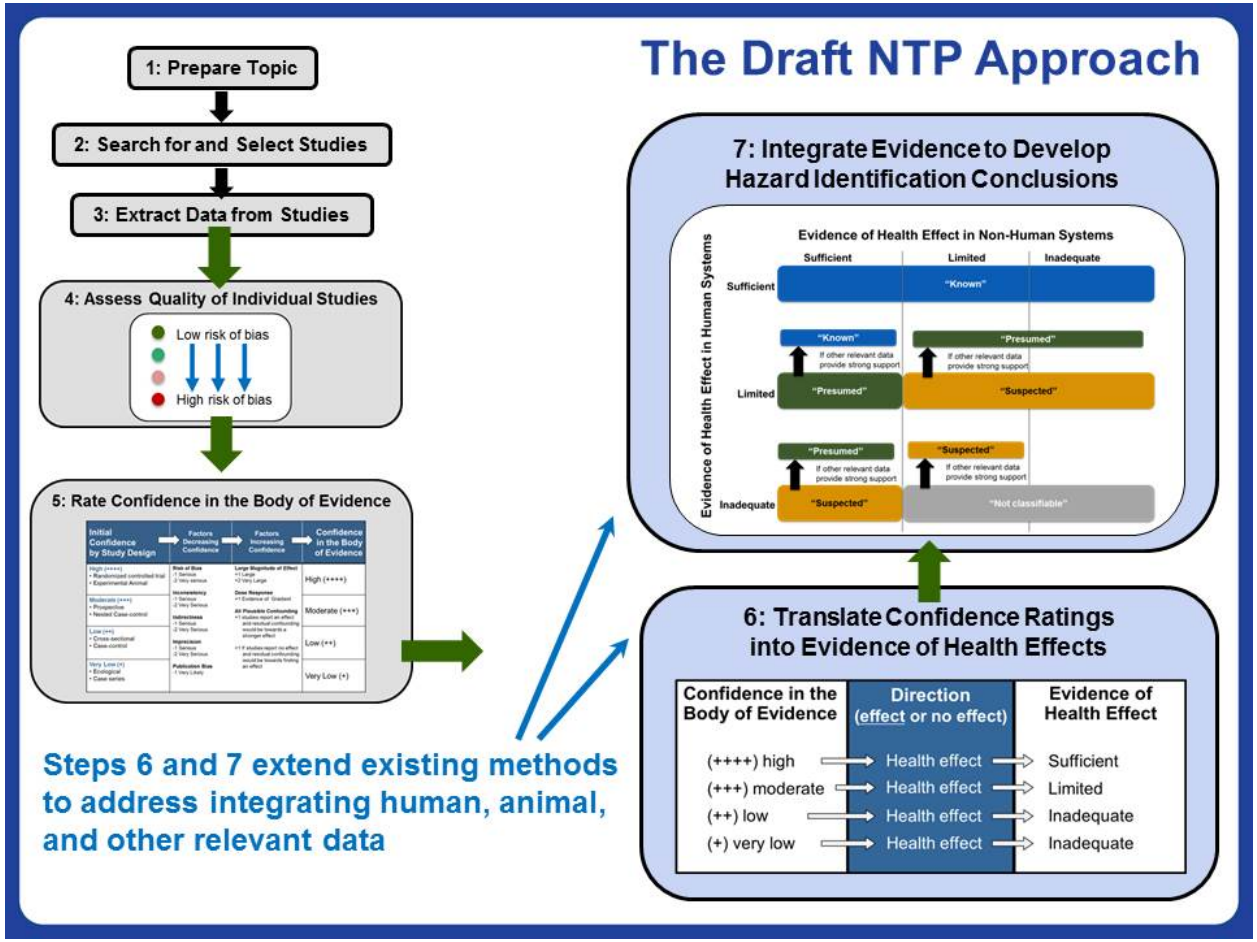

The Draft NTP Approach
1: Prepare Topic
2: Search for and Select Studies
3: Extract Data from Studies
4: Assess Quality of Individual Studies
Low risk of bias
High risk of bias
5: Rate Confidence in the Body of Evidence
Initial Confidence by Study Design
Factors Decreasing Confidence
Factors Increasing Confidence
Confidence in the Body of Evidence
High (++++)
• Randomized controlled trial
• Experimental Animal
Moderate (+++)
• Prospective
• Nested Case-control
Low (++)
• Cross-sectional
• Case-control
Very Low (+)
• Ecological
• Case series
Risk of Bias
-1 Serious
-2 Very serious
Inconsistency
-1 Serious
-2 Very Serious
Indirectness
-1 Serious
-2 Very Serious
Imprecision
-1 Serious
-2 Very Serious
Publication Bias
-1 Very Likely
Large Magnitude of Effect
+1 Large
+2 Very Large
Dose Response
+1 Evidence of Gradient
All Plausible Confounding
+1 studies report an effect and residual confounding would be towards a stronger effect
+1 If studies report no effect and residual confounding would be towards finding an effect
High (++++)
Moderate (+++)
Low (++)
Very Low (+)
6: Translate Confidence Ratings into Evidence of Health Effects
Confidence in the Body of Evidence
Direction (effect or no effect)
Evidence of Health Effect
(++++) high
(+++) moderate
(++) low
(+) very low
Health effect
Sufficient
Limited
Inadequate
7: Integrate Evidence to Develop Hazard Identification Conclusions
Evidence of Health Effect in Non-Human Systems
Evidence of Health Effect in Human Systems
Sufficient
Limited
Inadequate
"Known"
"Presumed"
"Suspected"
"Not classifiable"
If other relevant data provide strong support

Presentation Outline

- Background on systematic review
- Development of the draft NTP Approach
- The draft NTP Approach and evidence integration
- **Specific aspects brought to working group for comment**
  - Step 4: Assessing the quality or risk of bias of individual studies
  - Step 5: Rating the confidence in the body of evidence
  - Step 6: Translating confidence ratings into evidence of health effects
  - Step 7: Integrating evidence to develop hazard identification conclusions
- Questions


The Draft NTP Approach
1: Prepare Topic
2: Search for and Select Studies
3: Extract Data from Studies
First steps (1-3) are essentially the same as existing methods
4: Assess Quality of Individual Studies
Low risk of bias
High risk of bias
5: Rate Confidence in the Body of Evidence
Initial Confidence by Study Design
Factors Decreasing Confidence
Factors Increasing Confidence
Confidence in the Body of Evidence
High (++++)
• Randomized controlled trial
• Experimental Animal
Moderate (+++)
• Prospective
• Nested Case-control
Low (++)
• Cross-sectional
• Case-control
Very Low (+)
• Ecological
• Case series
Risk of Bias
-1 Serious
-2 Very serious
Inconsistency
-1 Serious
-2 Very Serious
Indirectness
-1 Serious
-2 Very Serious
Imprecision
-1 Serious
-2 Very Serious
Publication Bias
-1 Very Likely
Large Magnitude of Effect
+1 Large
+2 Very Large
Dose Response
+1 Evidence of Gradient
All Plausible Confounding
+1 studies report an effect and residual confounding would be towards a stronger effect
+1 If studies report no effect and residual confounding would be towards finding an effect
High (++++)
Moderate (+++)
Low (++)
Very Low (+)
6: Translate Confidence Ratings into Evidence of Health Effects
Confidence in the Body of Evidence
Direction (effect or no effect)
Evidence of Health Effect
(++++) high
(+++) moderate
(++) low
(+) very low
Health effect
Sufficient
Limited
Inadequate
7: Integrate Evidence to Develop Hazard Identification Conclusions
Evidence of Health Effect in Non-Human Systems
Sufficient
Limited
Inadequate
"Known"
"Presumed"
"Suspected"


The Draft NTP Approach
1: Prepare Topic
2: Search for and Select Studies
3: Extract Data from Studies
Steps 4 and 5 build on existing methods with adaptations to address the types of data relevant for environmental health questions
4: Assess Quality of Individual Studies
Low risk of bias
High risk of bias
5: Rate Confidence in the Body of Evidence
Initial Confidence by Study Design
Factors Decreasing Confidence
Factors Increasing Confidence
Confidence in the Body of Evidence
High (++++)
• Randomized controlled trial
• Experimental Animal
Moderate (+++)
• Prospective
• Nested Case-control
Low (++)
• Cross-sectional
• Case-control
Very Low (+)
• Ecological
• Case series
Risk of Bias
-1 Serious
-2 Very serious
Inconsistency
-1 Serious
-2 Very Serious
Indirectness
-1 Serious
-2 Very Serious
Imprecision
-1 Serious
-2 Very Serious
Publication Bias
-1 Very Likely
Large Magnitude of Effect
+1 Large
+2 Very Large
Dose Response
+1 Evidence of Gradient
All Plausible Confounding
+1 studies report an effect and residual confounding would be towards a stronger effect
+1 If studies report no effect and residual confounding would be towards finding an effect
High (++++)
Moderate (+++)
Low (++)
Very Low (+)
6: Translate Confidence Ratings into Evidence of Health Effects
Confidence in the Body of Evidence
Direction (effect or no effect)
Evidence of Health Effect
(++++) high
(+++) moderate
(++) low
(+) very low
Health effect
Sufficient
Limited
Inadequate
7: Integrate Evidence to Develop Hazard Identification Conclusions
Evidence of Health Effect in Non-Human Systems
Sufficient
Limited
Inadequate
"Known"
"Presumed"
"Suspected"


The Draft NTP Approach
1: Prepare Topic
2: Search for and Select Studies
3: Extract Data from Studies
4: Assess Quality of Individual Studies
Low risk of bias
High risk of bias
5: Rate Confidence in the Body of Evidence
Initial Confidence by Study Design
Factors Decreasing Confidence
Factors Increasing Confidence
Confidence in the Body of Evidence
High (++++)
Moderate (+++)
Low (++)
Very Low (+)
7: Integrate Evidence to Develop Hazard Identification Conclusions
Evidence of Health Effect in Non-Human Systems
Evidence of Health Effect in Human Systems
Sufficient
Limited
Inadequate
"Known"
"Presumed"
"Suspected"
"Not classifiable"
If other relevant data provide strong support
6: Translate Confidence Ratings into Evidence of Health Effects
Confidence in the Body of Evidence
Direction (effect or no effect)
Evidence of Health Effect
(++++) high
(+++) moderate
(++) low
(+) very low
Health effect
Sufficient
Limited
Inadequate
Steps 6 and 7 extend existing methods to address integrating human, animal, and other relevant data

NTP BSC Working Group

- NTP BSC Working Group members
  - **Lynn Goldman -** Chair, Dean and Professor, George Washington University
  - **Reeder Sams -** Vice-Chair, Acting Deputy Director, National Center for Environmental Assessment/RTP Div., USEPA
  - **Lisa Bero -** Director, Cochrane Center at UCSF
  - **Edward Carney -** Senior Science Leader, Mammalian Toxicology, Dow Chemical Company
  - **David Dorman -** Professor, North Carolina State University
  - **Elaine Faustman –** Director, Institute for Risk Analysis and Risk Communication, University of Washington
  - **Dale Hattis -** Research Professor, George Perkins Marsh Institute, Clark University
  - **Malcolm Macleod -** CAMARADES Centre, University of Edinburgh
  - **Tracey Woodruff -** Director, Program on Reproductive Health and the Environment, UCSF
  - **Lauren Zeise –** Chief, Reproductive and Cancer Hazard Assessment Branch, OEHHA, California EPA

- Meeting on August 28-29 in Raleigh, NC
  - **Charge**:

    to obtain feedback on the NTP's proposed approach for reaching conclusions for literature-based evidence assessments

  - **Goal**:

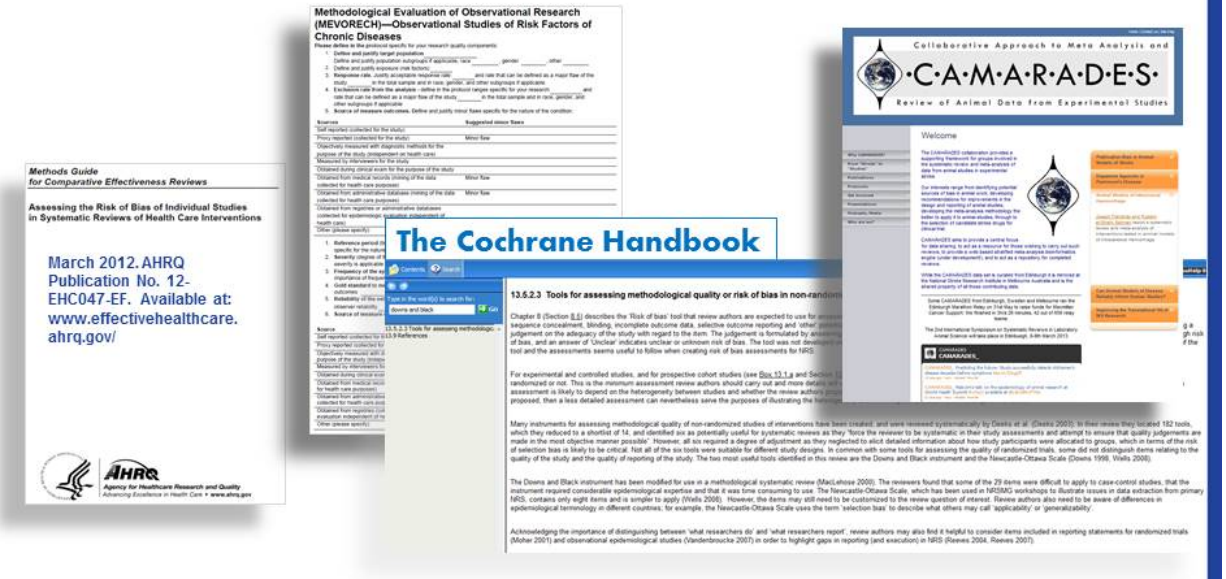
    *to get input on **specific aspects** of the draft NTP Approach*

Step 4: Assess the Quality of Individual Studies

- **Study quality or risk of bias**
  - Are you confident in the study findings?
- **Existing methods**
  - Established risk of bias tools for randomized controlled trials
  - Single summary scores for "study quality" are strongly discouraged
  - Reporting quality checklists **<u>are not</u>** risk of bias tools
  - No existing consensus on how to assess risk of bias for
    - Observational human studies, or
    - Animal studies

Adaptation of Existing Study Quality Methods

- Although there are a variety of risk of bias methods for human studies, animal tools are generally reporting quality checklists (e.g., ToxRTool)



* Assessing the Risk of Bias of Individual Studies in Systematic Reviews of Health Care Interventions (AHRQ, Viswanathan, 2012)

Adaptation of Existing Study Quality Methods

- Although there are a variety of risk of bias methods for human studies, animal tools are generally reporting quality checklists (e.g., ToxRTool)
- The recent AHRQ method guide* was particularly useful as a model because it covers RCTs and a range of human observational studies

Table 4. Design-specific criteria to assess for risk of bias for benefits

Risk of bias	Criterion	RCTs	CCTs or cohort	Case-control	Case series	Cross-sectional
Selection bias	Was the allocation sequence generated adequately (e.g., random number table, computer-generated randomization)?	x				
	Was the allocation of treatment adequately concealed (e.g., pharmacy- controlled randomization or use of sequentially numbered sealed envelopes)?	x				
	Were participants analyzed within the groups they were originally assigned to?	x	x			
	Did the study apply inclusion/exclusion criteria uniformly to all comparison groups?		x			x
	Were cases and controls selected appropriately (e.g., appropriate diagnostic criteria or definitions, equal application of exclusion criteria to case and controls, sampling not influenced by exposure status)			x		
	Did the strategy for recruiting participants into the study differ across study groups?		x			
	Does the design or analysis control account for important confounding and modifying variables through matching, stratification, multivariable analysis, or other approaches?	x	x	x	x	x
Performance bias	Did researchers rule out any impact from a concurrent intervention or an unintended exposure that might bias results?	x	x	x	x	x
	Did the study maintain fidelity to the intervention protocol?	x	x	x	x	
Attrition bias	If attrition (overall or differential nonresponse, dropout, loss to follow-up, or exclusion of participants) was a concern, were missing data handled appropriately (e.g., intention-to-treat analysis and imputation)?	x	x	x	x	x
Detection bias	In prospective studies, was the length of follow-up different between the groups, or in case-control studies, was the time period between the intervention/exposure and outcome the same for cases and controls?	x	x	x		
	Were the outcome assessors blinded to the intervention or exposure status of participants?	x	x	x	x	x
	Were interventions/exposures assessed/defined using valid and reliable measures, implemented consistently across all study participants?	x	x	x	x	x
	Were outcomes assessed/defined using valid and reliable measures, implemented consistently across all study participants?	x	x	x	x	x
	Were confounding variables assessed using valid and reliable measures, implemented consistently across all study participants?		x	x	x	x
Reporting bias	Were the potential outcomes prespecified by the researchers? Are all prespecified outcomes reported?	x	x	x	x	x

*Cases and controls should be similar in all factors known to be associated with the disease of interest, but they should not be so uniform as to be matched for the exposure of interest.

* Assessing the Risk of Bias of Individual Studies in Systematic Reviews of Health Care Interventions (AHRQ, Viswanathan, 2012)

Adaptation of Existing Study Quality Methods

- Although there are a variety of risk of bias methods for human studies, animal tools are generally reporting quality checklists (e.g., ToxRTool)
- The recent AHRQ method guide* was particularly useful as a model because it covers RCTs and a range of human observational studies

Consideration of 5 traditional risk of bias domains

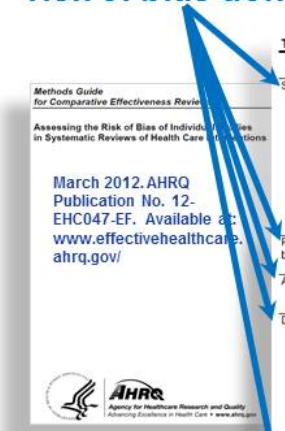


Table 4. Design-specific criteria to assess for risk of bias for benefits

Risk of bias	Criterion	RCTs	CCTs or cohort	Case-control	Case series	Cross-sectional
Selection bias	Was the allocation sequence generated adequately (e.g., random number table, computer-generated randomization)?	x				
	Was the allocation of treatment adequately concealed (e.g., pharmacy- controlled randomization or use of sequentially numbered sealed envelopes)?	x				
	Were participants analyzed within the groups they were originally assigned to?	x	x			
	Did the study apply inclusion/exclusion criteria uniformly to all comparison groups?		x			x
	Were cases and controls selected appropriately (e.g., appropriate diagnostic criteria or definitions, equal application of exclusion criteria to case and controls, sampling not influenced by exposure status)			x		
	Did the strategy for recruiting participants into the study differ across study groups?		x			
	Does the design or analysis control account for important confounding and modifying variables through matching, stratification, multivariable analysis, or other approaches?	x	x	x	x	x
Performance bias	Did researchers rule out any impact from a concurrent intervention or an unintended exposure that might bias results?	x	x	x	x	x
	Did the study maintain fidelity to the intervention protocol?	x	x	x	x	
Attrition bias	If attrition (overall or differential nonresponse, dropout, loss to follow-up, or exclusion of participants) was a concern, were missing data handled appropriately (e.g., intention-to-treat analysis and imputation)?	x	x	x	x	x
Detection bias	In prospective studies, was the length of follow-up different between the groups, or in case-control studies, was the time period between the intervention/exposure and outcome the same for cases and controls?	x	x	x		
	Were the outcome assessors blinded to the intervention or exposure status of participants?	x	x	x	x	x
	Were interventions/exposures assessed/defined using valid and reliable measures, implemented consistently across all study participants?	x	x	x	x	x
	Were outcomes assessed/defined using valid and reliable measures, implemented consistently across all study participants?	x	x	x	x	x
	Were confounding variables assessed using valid and reliable measures, implemented consistently across all study participants?		x	x	x	x
Reporting bias	Were the potential outcomes prespecified by the researchers? Are all prespecified outcomes reported?	x	x	x	x	x

*Cases and controls should be similar in all factors known to be associated with the disease of interest, but they should not be so uniform as to be matched for the exposure of interest.

* Assessing the Risk of Bias of Individual Studies in Systematic Reviews of Health Care Interventions (AHRQ, Viswanathan, 2012)

Adaptation of Existing Study Quality Methods

- Although there are a variety of risk of bias methods for human studies, animal tools are generally reporting quality checklists (e.g., ToxRTool)
- The recent AHRQ method guide* was particularly useful as a model because it covers RCTs and a range of human observational studies

Study design determines which questions apply

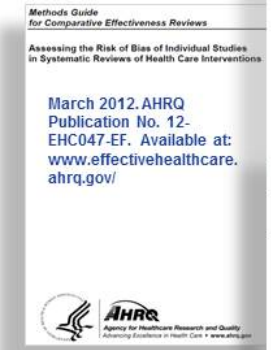


Table 4. Design-specific criteria to assess for risk of bias for benefits

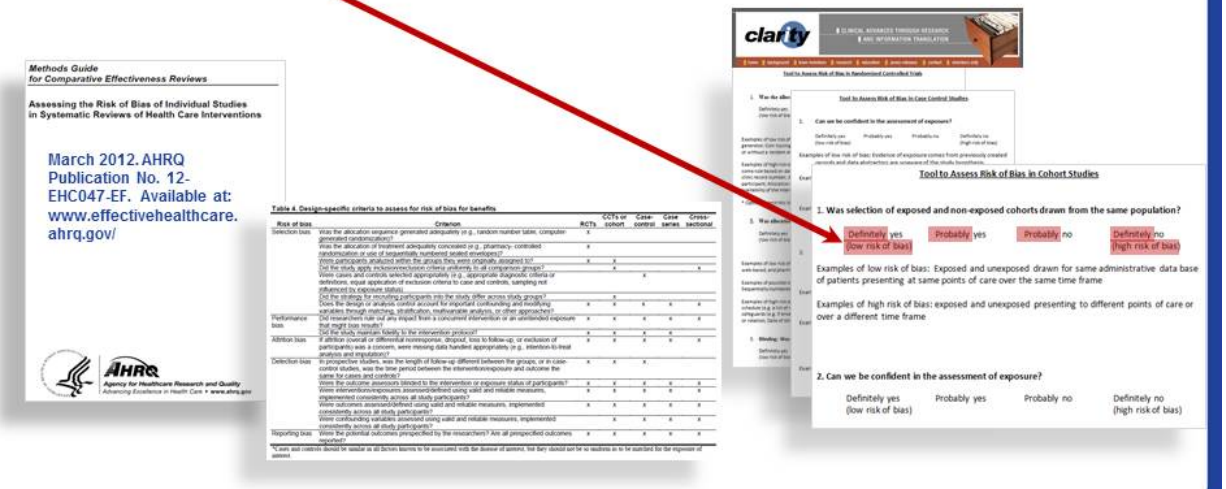
Risk of bias	Criterion	RCTs	CCTs or cohort	Case-control	Case series	Cross-sectional
Selection bias	Was the allocation sequence generated adequately (e.g., random number table, computer-generated randomization)?	x				
	Was the allocation of treatment adequately concealed (e.g., pharmacy- controlled randomization or use of sequentially numbered sealed envelopes)?	x				
	Were participants analyzed within the groups they were originally assigned to?	x	x			
	Did the study apply inclusion/exclusion criteria uniformly to all comparison groups?		x			x
	Were cases and controls selected appropriately (e.g., appropriate diagnostic criteria or definitions, equal application of exclusion criteria to case and controls, sampling not influenced by exposure status)			x		
	Did the strategy for recruiting participants into the study differ across study groups?		x			
	Does the design or analysis control account for important confounding and modifying variables through matching, stratification, multivariable analysis, or other approaches?	x	x	x	x	x
Performance bias	Did researchers rule out any impact from a concurrent intervention or an unintended exposure that might bias results?	x	x	x	x	x
	Did the study maintain fidelity to the intervention protocol?	x	x	x	x	
Attrition bias	If attrition (overall or differential nonresponse, dropout, loss to follow-up, or exclusion of participants) was a concern, were missing data handled appropriately (e.g., intention-to-treat analysis and imputation)?	x	x	x	x	x
Detection bias	In prospective studies, was the length of follow-up different between the groups, or in case-control studies, was the time period between the intervention/exposure and outcome the same for cases and controls?	x	x	x		
	Were the outcome assessors blinded to the intervention or exposure status of participants?	x	x	x	x	x
	Were interventions/exposures assessed/defined using valid and reliable measures, implemented consistently across all study participants?	x	x	x	x	x
	Were outcomes assessed/defined using valid and reliable measures, implemented consistently across all study participants?	x	x	x	x	x
	Were confounding variables assessed using valid and reliable measures, implemented consistently across all study participants?		x	x	x	x
Reporting bias	Were the potential outcomes prespecified by the researchers? Are all prespecified outcomes reported?	x	x	x	x	x

*Cases and controls should be similar in all factors known to be associated with the disease of interest, but they should not be so uniform as to be matched for the exposure of interest.

* Assessing the Risk of Bias of Individual Studies in Systematic Reviews of Health Care Interventions (AHRQ, Viswanathan, 2012)

Adaptation of Existing Study Quality Methods

- Although there are a variety of risk of bias methods for human studies, animal tools are generally reporting quality checklists (e.g., ToxRTool)
- The recent AHRQ method guide* was particularly useful as a model because it covers RCTs and a range of human observational studies
- The clarity group scale for answering risk of bias questions was also useful (definitely low, probably low, probably high, to definitely high)



* Assessing the Risk of Bias of Individual Studies in Systematic Reviews of Health Care Interventions (AHRQ, Viswanathan, 2012)

The NTP Method to Assess Quality or Risk of Bias of Individual Studies

- Judge whether the design and conduct of individual studies compromise credibility of the link between exposure and outcome
- Evaluation is endpoint/outcome specific
- **Major issues brought to the BSC working group (WG) for comment**
  - Study quality evaluated with set of risk of bias questions based on AHRQ
  - Same questions adapted to also address experimental animal studies
  - Risk of bias answers from clarity group (definitely low, probably low, probably high, definitely high)
  - Proposed "Major" risk of bias questions as having greater impact on confidence that environmental substances are associated with health effects (e.g., *"Can we be confident in the exposure assessment?"*)

Step 5: Rate Confidence in the Body of Evidence

- **Confidence Rating**
  - How confident are you that findings from a group of studies reflect the true relationship between exposure to a substance and an effect?
- **Existing Methods**
  - The GRADE approach is a widely accepted method for rating confidence in a body of evidence
    - No guidance for animal studies
    - All observational human studies are given the same initial low quality (e.g., case-report = prospective cohort study)

Why GRADE?

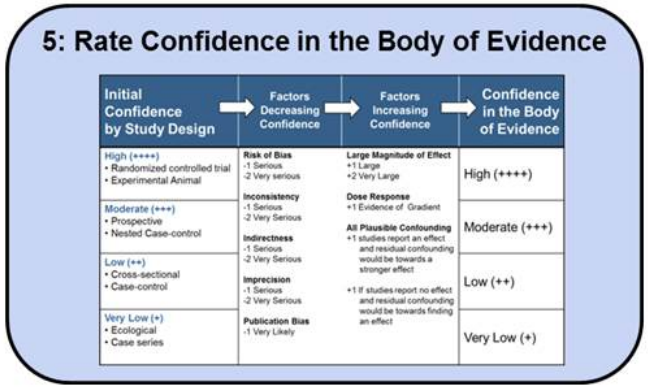
- Developed by broad group of international guideline developers in the area of healthcare
- Clear presentation of elements considered for downgrading or upgrading confidence in body of evidence
  - Framework for documenting scientific judgment decisions
  - Elements cover Bradford Hill criteria
  - Practitioners engage in ongoing methods development

Why GRADE?

- Developed by broad group of international guideline developers in the area of healthcare
- Clear presentation of elements considered for downgrading or upgrading confidence in body of evidence
  - Framework for documenting scientific judgment decisions
  - Elements cover Bradford Hill criteria
  - Practitioners engage in ongoing methods development
- Endorsed and used by over 70 organizations
- Consistent with DHHS sister agencies
  - Conceptually similar to AHRQ model
  - Supported by parts of CDC for healthcare recommendations

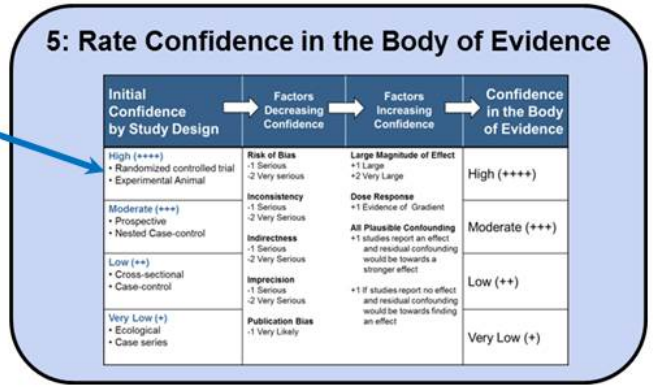
The NTP Method to Rate Confidence in the Body of Evidence

- Rate confidence that findings from a group of studies reflect the true relationship between exposure to a substance and an effect
- **Major issues brought to BSC WG for comment**
  - Method for rating confidence based on GRADE and AHRQ approaches adapted to address data relevant for environmental health questions



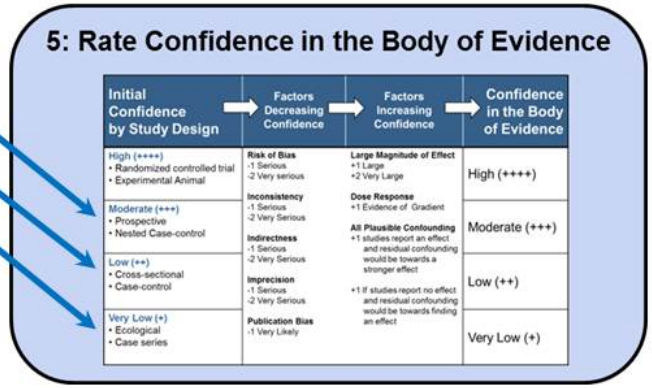
The NTP Method to Rate Confidence in the Body of Evidence

- Rate confidence that findings from a group of studies reflect the true relationship between exposure to a substance and an effect
- **Major issues brought to BSC WG for comment**
  - Method for rating confidence based on GRADE and AHRQ approaches adapted to address data relevant for environmental health questions
  - **Initial confidence based on study design**
    - Experimental animal studies at same initial rating as RCTs



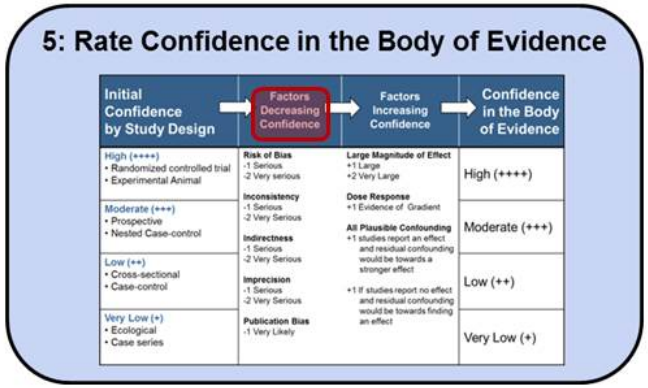
The NTP Method to Rate Confidence in the Body of Evidence

- Rate confidence that findings from a group of studies reflect the true relationship between exposure to a substance and an effect
- **Major issues brought to BSC WG for comment**
  - Method for rating confidence based on GRADE and AHRQ approaches adapted to address data relevant for environmental health questions
  - **Initial confidence based on study design**
    - Experimental animal studies at same initial rating as RCTs
    - Broader initial confidence rating to address range of human observational studies



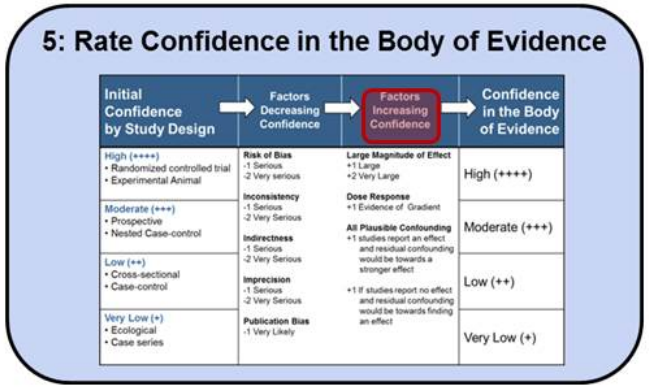
The NTP Method to Rate Confidence in the Body of Evidence

- Rate confidence that findings from a group of studies reflect the true relationship between exposure to a substance and an effect
- **Major issues brought to BSC WG for comment**
  - Method for rating confidence based on GRADE and AHRQ approaches adapted to address data relevant for environmental health questions
  - **Initial confidence based on study design**
    - Experimental animal studies at same initial rating as RCTs
    - Broader initial confidence rating to address range of human observational studies
  - **Decreasing/Increasing**
    - Factors for decreasing confidence consistent with GRADE approach



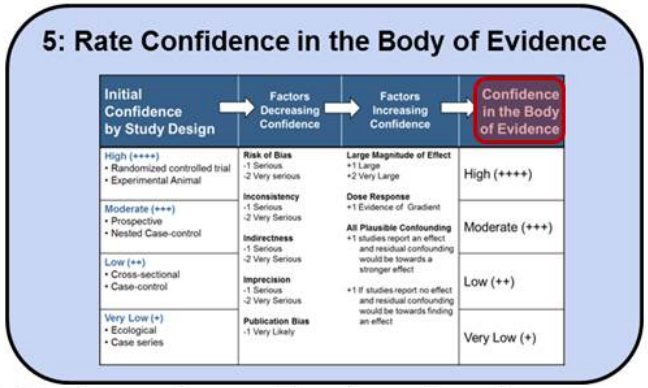
The NTP Method to Rate Confidence in the Body of Evidence

- Rate confidence that findings from a group of studies reflect the true relationship between exposure to a substance and an effect
- **Major issues brought to BSC WG for comment**
  - Method for rating confidence based on GRADE and AHRQ approaches adapted to address data relevant for environmental health questions
  - **Initial confidence based on study design**
    - Experimental animal studies at same initial rating as RCTs
    - Broader initial confidence rating to address range of human observational studies
  - **Decreasing/Increasing**
    - Additional factors considered for increasing confidence (e.g., consistency across animal models or species)



The NTP Method to Rate Confidence in the Body of Evidence

- Rate confidence that findings from a group of studies reflect the true relationship between exposure to a substance and an effect
- **Major issues brought to BSC WG for comment**
  - Method for rating confidence based on GRADE and AHRQ approaches adapted to address data relevant for environmental health questions
  - **Initial confidence based on study design**
    - Experimental animal studies at same initial rating as RCTs
    - Broader initial confidence rating to address range of human observational studies
  - **Decreasing/Increasing**
    - Additional factors considered for increasing confidence (e.g., consistency across animal models or species)
  - **Confidence rating by endpoint/outcome is used in steps 6 and 7**

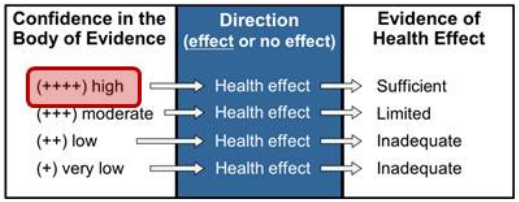


Step 6: Translate Confidence Ratings into Level of Evidence for Health Effects

- **Level of Evidence**
  - What is the level of evidence for a health effect (or no effect)?
- **Additional step is necessary to consider both**
  - Confidence in the association between exposure and outcome, and
  - Direction of the effect (toxicity or no toxicity)

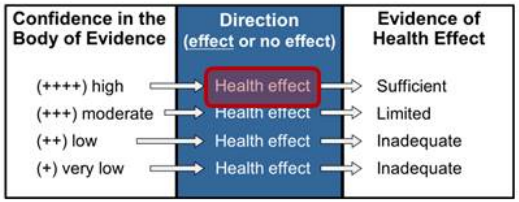
Step 6: Translate Confidence Ratings into Level of Evidence for Health Effects

- **Level of Evidence**
  - What is the level of evidence for a health effect (or no effect)?
- **Additional step is necessary to consider both**
  - Confidence in the association between exposure and outcome, and
  - Direction of the effect (toxicity or no toxicity)



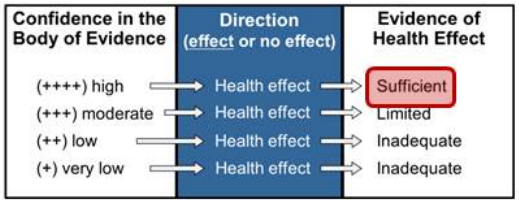
Step 6: Translate Confidence Ratings into Level of Evidence for Health Effects

- **Level of Evidence**
  - What is the level of evidence for a health effect (or no effect)?
- **Additional step is necessary to consider both**
  - Confidence in the association between exposure and outcome, and
  - Direction of the effect (toxicity or no toxicity)



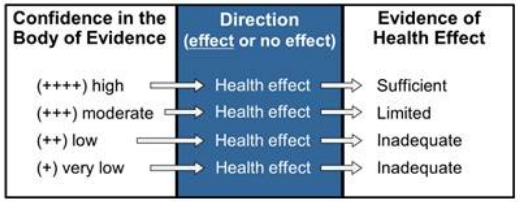
Step 6: Translate Confidence Ratings into Level of Evidence for Health Effects

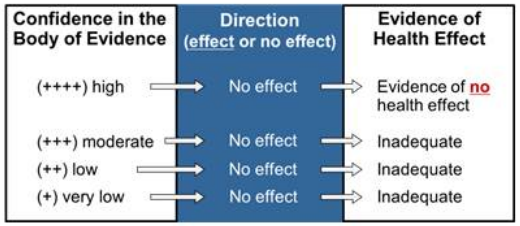
- **Level of Evidence**
  - What is the level of evidence for a health effect (or no effect)?
- **Additional step is necessary to consider both**
  - Confidence in the association between exposure and outcome, and
  - Direction of the effect (toxicity or no toxicity)



Step 6: Translate Confidence Ratings into Level of Evidence for Health Effects

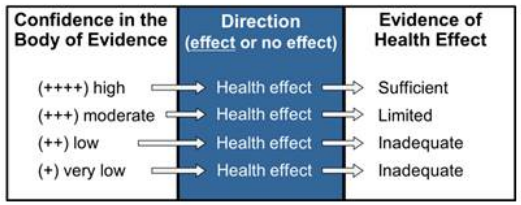
- **Level of Evidence**
  - What is the level of evidence for a health effect (or no effect)?
- **Additional step is necessary to consider both**
  - Confidence in the association between exposure and outcome, and
  - Direction of the effect (toxicity or no toxicity)

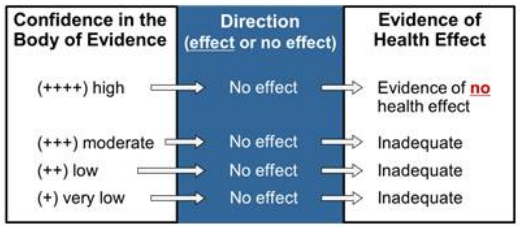




Step 6: Translate Confidence Ratings into Level of Evidence for Health Effects

- **Level of Evidence**
  - What is the level of evidence for a health effect (or no effect)?
- **Additional step is necessary to consider both**
  - Confidence in the association between exposure and outcome, and
  - Direction of the effect (toxicity or no toxicity)
- **Major issues brought to BSC WG for comment**
  - Evidence of health effects can be either "sufficient", "limited", or "inadequate"
  - A conclusion of evidence of no health effect requires high confidence



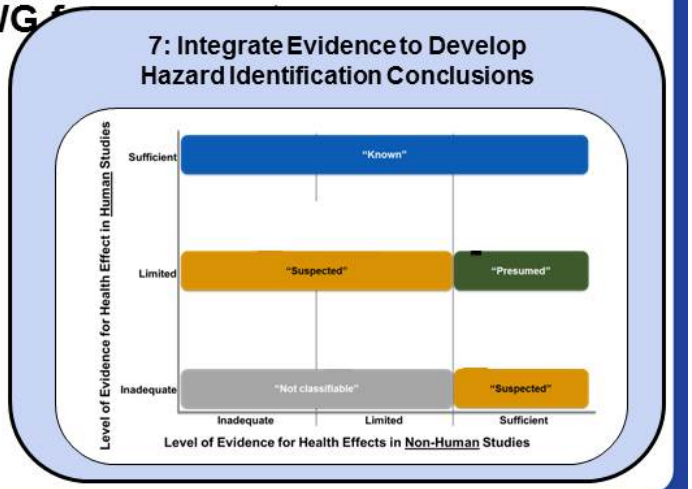


Step 7: Integrate Evidence to Develop Hazard Identification Conclusions

- **Integrate the Evidence**
  - What hazard ID conclusion is supported by considering the human, animal, and other relevant data together?
- **Additional step to integrate evidence and reach a conclusion**
  - Known, Presumed, Suspected, or Not classifiable to be a hazard to humans
- **Major issues brought to WG for comment**
  - Two part process to combine evidence streams

Step 7: Integrate Evidence to Develop Hazard Identification Conclusions

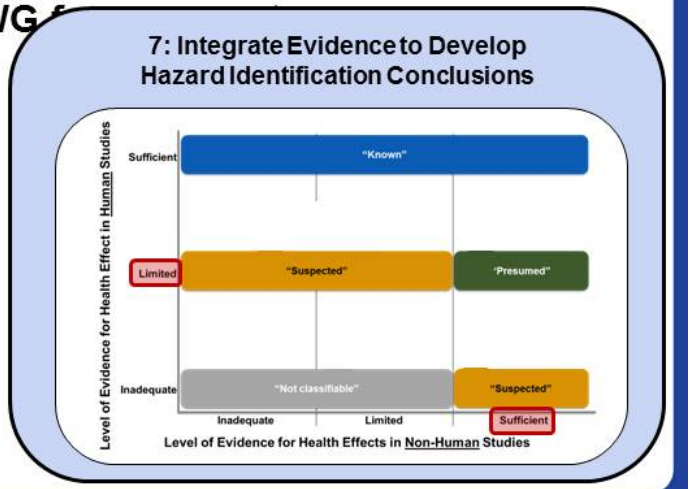
- **Integrate the Evidence**
  - What hazard ID conclusion is supported by considering the human, animal, and other relevant data together?
- **Additional step to integrate evidence and reach a conclusion**
  - Known, Presumed, Suspected, or Not classifiable to be a hazard to humans
- **Major issues brought to WG f**
  - Two part process to combine evidence streams
    - **First**: human x animal



Step 7: Integrate Evidence to Develop Hazard Identification Conclusions

- **Integrate the Evidence**
  - What hazard ID conclusion is supported by considering the human, animal, and other relevant data together?
- **Additional step to integrate evidence and reach a conclusion**
  - Known, Presumed, Suspected, or Not classifiable to be a hazard to humans
- **Major issues brought to WG f**
  - Two part process to combine evidence streams
    - **First**: human x animal

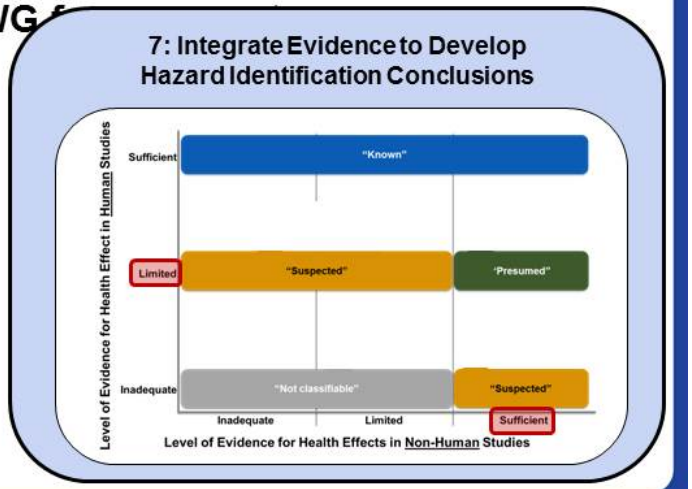
Consideration of animal data can increase hazard ID conclusion from human alone (if human evidence is Limited or Inadequate)



Step 7: Integrate Evidence to Develop Hazard Identification Conclusions

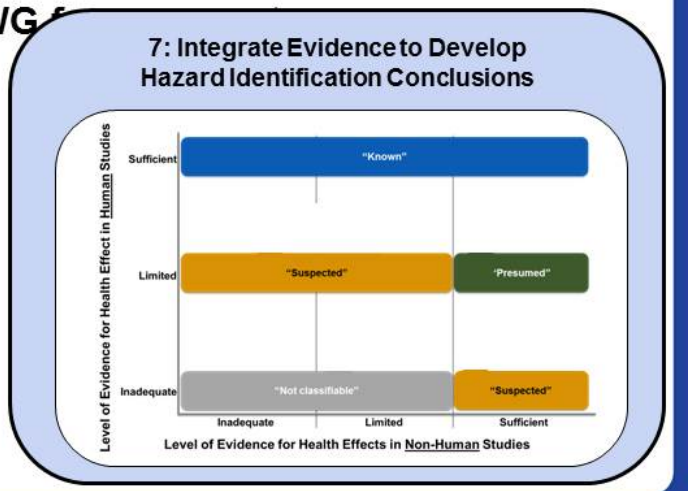
- **Integrate the Evidence**
  - What hazard ID conclusion is supported by considering the human, animal, and other relevant data together?
- **Additional step to integrate evidence and reach a conclusion**
  - Known, Presumed, Suspected, or Not classifiable to be a hazard to humans
- **Major issues brought to WG f**
  - Two part process to combine evidence streams
    - **First**: human x animal
    - **Second**: consider impact of other relevant data (e.g., mechanistic, *in vitro*, upstream indicator)

Consideration of other relevant data can increase hazard ID



Step 7: Integrate Evidence to Develop Hazard Identification Conclusions

- **Integrate the Evidence**
  - What hazard ID conclusion is supported by considering the human, animal, and other relevant data together?
- **Additional step to integrate evidence and reach a conclusion**
  - Known, Presumed, Suspected, or Not classifiable to be a hazard to humans
- **Major issues brought to WG f**
  - Two part process to combine evidence streams
    - **First**: human x animal
    - **Second**: consider impact of other relevant data (e.g., mechanistic, *in vitro*, upstream indicator)



Acknowledgements

- **Office of Health Assessment and Translation**
  - Abee Boyles
  - Kembra Howdeshell
  - Andrew Rooney, Deputy Director
  - Michael Shelby
  - Kyla Taylor
  - Kristina Thayer, Director
  - Vickie Walker
- **Office of Liaison, Policy and Review**
  - Mary Wolfe, Director
  - Lori White
- **Technical Advisors and Experts**
  - **Lisa Bero,** Director, San Francisco Branch, United States Cochrane Center at UC San Francisco
  - **Gordon Guyatt,** Co-chair, GRADE Working Group, McMaster University
  - **Malcolm Macleod,** CAMARADES Centre, University of Edinburgh
  - **Karen Robinson,** Co-Director, Evidence-Based Practice Center, The Johns Hopkins Bloomberg School of Public Health
  - **Holger Schünemann**, Co-chair, GRADE Working Group, McMaster University
  - **Tracey Woodruff,** Director, Program on Reproductive Health and the Environment, UCSF
- **NTP BSC Working Group**
  - **Lynn Goldman, Chair,** Dean, School of Public Health and Health Services, George Washington University, Washington, DC
  - **Reeder Sams, Vice-chair,** Acting Deputy Director, National Center for Environmental Assessment/RTP Division, USEPA
  - **Lisa Bero,** Director, San Francisco Branch, United States Cochrane Center at UC San Francisco
  - **Edward Carney,** Senior Science Leader, Mammalian Toxicology, Dow Chemical Company
  - **David Dorman,** Professor, North Carolina State University
  - **Elaine Faustman**, Director Institute for Risk Analysis and Risk Communication, University of Washington
  - **Dale Hattis**, Research Professor, George Perkins Marsh Institute, Clark University
  - **Malcolm Macleod,** CAMARADES Centre, University of Edinburgh
  - **Tracey Woodruff,** Director, Program on Reproductive Health and the Environment, UCSF
  - **Lauren Zeise,** Chief, Reproductive and Cancer Hazard Assessment Branch, OEHHA, California EPA

Questions?